

# Endpoints for Mouse Abdominal Tumor Models: Refinement of Current Criteria

Eden V Paster,<sup>1</sup> Kimberly A Villines,<sup>2</sup> and Debra L Hickman<sup>3,\*</sup>

Accurate, rapid, and noninvasive health assessments are required to establish more appropriate endpoints in mouse cancer models where tumor size is not easily measured. We evaluated potential endpoints in mice with experimentally induced peritoneal lymphoma, an abdominal tumor model, by comparing body weight, body condition, and behavior with those of a control group of mice not developing lymphoma. Our hypothesis was that body weight would increase or plateau, whereas body condition and behavioral scores would decrease, as disease progressed. Results indicated that body weight did not differ significantly between the control and experimental groups, but the experimental group experienced significant decreases in both body condition and behavioral scores. Our results support the use of body condition and behavioral scoring as adjunctive assessment methods for mice involved in abdominal lymphoma tumor studies in which health may decline despite an increase or plateau in body weight.

**Abbreviation:** BCS, body condition score.

Currently many approaches are used to monitor morbidity in experimental mouse studies, including assessment of body weight, activity, hydration, and hair coat appearance, with assessment of body weight predominating.<sup>1,11,15–17,19,21,23,25</sup> However, in cancer research, these commonly used endpoints are not always effective tools for health assessment. The assessment of activity or coat appearance is subjective. Dehydration and weight loss can be difficult to ascertain accurately because an increase in tumor mass can mask the loss of overall body weight that is associated with dehydration, loss of normal fat deposits, and muscle wasting. In mice with subcutaneous tumors, tumor size, tumor ulceration, and the animal's ability to ambulate can be measured objectively and used to evaluate health.<sup>8</sup> However, in mice with internal tumors, these parameters may be difficult to evaluate, and body weight and overall appearance may be the only parameters that can be assessed clearly. Therefore, additional evaluation methods that are noninvasive, reliable, and easily performed would be useful.

Body condition scoring (BCS) is a routinely used technique in veterinary medicine for assessing health and nutritional status. In ruminants, pigs, and horses, BCS has been used to assess health in disease and reproductive states.<sup>3,4,9,14,18,22</sup> In addition, BCS techniques have been used to monitor dogs and cats with neoplasia and heart disease<sup>2,20</sup> and to evaluate diet choice and volume when treating obesity.<sup>12,13</sup> Furthermore, BCS techniques have been developed for application to laboratory species and have the potential to improve animal welfare in research.<sup>5,6,7,8,10,24</sup>

Body condition scoring has been adapted for rodents. In rats, BCS and body weight have been used adjunctively to evaluate diabetes models.<sup>11</sup> Although BCS techniques for mice have been

used to accurately assess the health of P- and E-selectin double-deficient mice,<sup>7,24</sup> this technique has not been applied to or evaluated in other mouse models.<sup>6,24</sup> In addition, although BCS is an effective evaluation method, it alone does not give a complete picture of animal health. In this study, we used body weight, BCS, appearance, and behavioral assessments to evaluate morbidity in a mouse model of peritoneal lymphoma. Our hypothesis was that body weight would plateau or increase as the tumors increased in size, but body condition score would decrease and, therefore, more accurately reflect the true health status of the mouse. We also hypothesized that a change in appearance and behavior would accompany the decrease in BCS. A total score combining these evaluations was developed to assess overall health status in mice, helping investigators and animal care staff to reevaluate study endpoints for abdominal tumor growth. A further hypothesis was that the total score would provide more sensitivity than BCS, appearance, or behavior alone in assessing health status. To reduce the overall number of animals used for this study, we collaborated with an investigator performing abdominal lymphoma research at our institution and evaluated animals from ongoing studies of different anticancer vaccine therapies.

## Materials and Methods

**Animals.** Female C57BL/6J mice (n = 40; age, 7 wk; Jackson Laboratory, Bar Harbor, ME) were evaluated for this study. These mice were already being used in a collaborator's abdominal lymphoma study, which was designed by another investigator. The mice were each assigned a unique number and were housed together in groups of 4, according to our collaborator's treatment groups. The endpoint assessment control groups (C1 and C2) comprised a total of 16 mice, and the endpoint assessment experimental groups (E1, E2, and E3) involved 24 mice in all. Mice were housed in polycarbonate shoebox cages with filter tops (Thoren Caging Systems, Hazelton, PA) and corncob bedding (Bed-O'Cobs, The Andersons, Maumee, OH). Cages were

Received: 08 Aug 2008. Revision requested: 24 Sep 2008. Accepted: 11 Dec 2008.

<sup>1</sup>Veterinary Medical Unit, VA Medical Center, Portland, Oregon; <sup>2</sup>College of Veterinary Medicine, Oregon State University, Corvallis, Oregon; <sup>3</sup>Laboratory Animal Resource Center, Indiana University School of Medicine, Indianapolis, Indiana.

\*Corresponding author. Email: hickmand@iupui.edu

changed weekly in a laminar-flow changing station (Lab Products, Seaford, DE). Animal caretakers wore gloves and sprayed their gloves with 10% bleach solution between cage changes. Soiled cages were sanitized in a mechanical cage washer with a final rinse temperature of 82 °C. The room was kept on a 12:12-h light:dark cycle. Mice were provided with ad libitum rodent chow (LabDiet 5001, Purina Mills International, St Louis, MO) and tap water. The temperature in the room was maintained at 22 °C and humidity at 30% to 70%. All procedures were approved by the Portland VA Medical Center Institutional Animal Care and Use Committee in accordance with federal policy.

At the time of this study, all mice were free of mouse coronavirus, Sendai virus, mouse parvovirus, minute virus of mice, *Ectromelia* virus, reovirus type 3, pneumonia virus of mice, murine adenovirus, *Mycoplasma pulmonis*, lymphocytic choriomeningitis virus, mouse rotavirus, mouse encephalomyelitis virus, polyoma virus, murine cytomegalovirus, and rodent pinworms and mites as assessed through the quarterly evaluation of sentinels indirectly exposed to colony animal dirty bedding. Sentinel animals at this facility are not screened routinely for other pathogens such as *Helicobacter* spp. and murine norovirus.

**Endpoint assessment experimental and control groups.** In our collaborator's study, the mice were divided into control and experimental groups. The endpoint assessment control groups ( $n = 16$ ) consisted of mice that were not challenged with tumor cells (Figure 1). According to our collaborator's protocol, 24 mice were evaluated as part of the endpoint assessment experimental group. All of the animals in the experimental groups were challenged with the C6VL tumor line, as described in the next section.

**Induction of T-cell lymphoma and vaccination.** Mice in the endpoint assessment experimental group were inoculated with tumor cells followed by immunization with idiotypic protein vaccines with or without antibodies to assess the effectiveness of antitumor activity in our collaborator's experiment. The tumor cell line used in these experiments was C6VL, which is a thymoma induced by radiation in C57BL/6J mice. The tumor cell line originated inhouse from mice that were screened for pathogens by our sentinel program. Cells for tumor induction were grown in vitro, frozen in liquid nitrogen, and thawed 1 to 2 d prior to tumor challenge. Tumor cells were washed in PBS, and a dose of 500  $\mu$ L containing 5000 cells was given intraperitoneally. The endpoint assessment control groups were injected intraperitoneally with 500  $\mu$ L PBS instead of being challenged with tumor cells. The vaccine was made from a unique malignant lymphocyte protein (idiotypic protein) mixed with various adjuvants. Idiotypic proteins were purified, dialyzed, and sterilized prior to administration. Vaccines (0.35 mg) were given in a volume of 0.1 to 0.2 mL either subcutaneously or intraperitoneally and were repeated every 1 to 2 wk for 3 to 4 immunizations. Control mice received PBS-only injections.

Some mice were treated with antiCD4 cell markers to test binding to the tumor variable region. These antibody treatments (250  $\mu$ g) were given in a volume of 500  $\mu$ L subcutaneously or intraperitoneally once weekly for 4 wk. The tumors grew quickly, and signs of abdominal distention were evident at 4 to 6 wk. Animals were checked by our collaborator and animal care staff daily for signs of morbidity. Consistent with the available literature, the institutional animal care and use committee approved endpoints that were based on indications of significant tumor growth, including 1 or a combination of these clinical signs: low BCS (that is,

1 on a scale of 1 to 5), abdominal distention that impeded movement, labored breathing, unkempt haircoat, abnormal posture, dehydration, and weight loss. If these clinical signs were observed by the technical staff, a veterinarian was consulted or the animal was euthanized. At 46 d after tumor induction or when IACUC-approved endpoints were achieved (regardless of the total score recorded), mice were euthanized by carbon dioxide asphyxiation. All but 1 animal in the experimental group were euthanized at various times before the 46-d study end due to achievement of euthanasia endpoints.

**Monitoring.** In addition to daily assessment by our collaborator, mice were assessed at the same time of the morning every other day by a group of 2 or 3 observers employed by our lab (instead of our collaborator) who evaluated each mouse against defined criteria (Figure 2). This assessment was performed to assign an objective rating to the subjective criteria that had been approved as part of the experimental protocol and to develop a panel of objective ratings that could be used to guide decisions regarding euthanasia of individual animals with improved consideration of animal welfare. The observers were blinded with regard to treatment by vaccine but performed their assessment concurrently. Our laboratory has demonstrated that there is no significant interobserver variation when assessing BCS in mice.<sup>8</sup> In addition, observers were blinded to endpoint assessment control and treatments groups, and a new score sheet was used daily to blind the observers to the previous days' data.

**Training of observers.** To obtain consistent scoring, our observer group received instruction and hands-on training in body condition, appearance, and behavior scoring, based on the criteria described in the following sections. Our lab practiced the scoring techniques together, and then separately, until the scoring system for each observer was reproducible and consistent with those of other observers. Our collaborator and his research staff completed online and instructor-led rodent handling and technique courses, as required by our institution. Our collaborator's group did not receive specific training in outscoring techniques, because scoring was done by our group.

**Appearance.** The filter top was removed, and each mouse was scored regarding its appearance. Animals were assigned a score of 2 to 0. A 'normal' score of 2 was given to mice having a shiny, well-kept hair coat; long, twitching vibrissae; bright, clear eyes; erect ears; and pink mucous membranes. A score of 1 was assigned to animals whose hair coat was unkempt, dull, or soiled and whose vibrissae appeared clumped. A score of 0 was given to an animal that was hunched; had bristled, clumped, soiled, dull fur; dry or dull eyes and nose; and tacky mucous membranes. Scores of 1 and 0 were considered abnormal.

**Natural behavior.** Before being handled, each mouse was assigned a score of 3 to 0 with regard to its unprovoked behavior ('natural behavior'). A normal mouse, receiving a score of 3, ambulated easily about the cage, took interest in its environment, interacted with its cagemates, and looked up at the observer. A mouse assigned a score of 2 took less interest in its environment, interacted less with cagemates, and disregarded the observer. A mouse receiving a score of 1 was less mobile, isolated from its cagemates, sat in a cage corner, and did not readily move when the cage was disturbed. A score of 0 was assigned if the mouse was immobile or hyperreactive. Scores of 0 and 1 were considered abnormal.

Control group (C): 16 mice total	C1 (4 mice)	Intraperitoneal PBS
	C2 (12 mice)	Intraperitoneal injection of PBS + antiVB6 chimeric tumor antibody
Experimental group (E): 24 mice total	E1 (12 mice)	Challenged with the C6VL tumor line + antiVB6 chimeric tumor antibody
	E2 (8 mice)	Challenged with the C6VL tumor line + antiCD4 chimeric tumor antibody along with an idiotype vaccine
	E3 (4 mice)	Challenged with the C6VL tumor line + antiVB7 chimeric tumor antibody

Figure 1. Description of control and experimental treatment groups.

Parameter	Description	Score
Appearance (Also note if abdominal distention is present)	N: bright eyes; shiny, well-groomed hair coat	2
	Abn: Unkempt hair coat, dull fur	1
	Abn: Hunching, piloerection	0
Natural behavior	N: Active; interactive in environment	3
	Slight decrease in activity; less interactive	2
	Abn: Pronounced decrease in activity; isolated	1
	Abn: Possible selfmutilation; hyperactive or immobile	0
Provoked behavior	N: Quickly moves away	3
	Slow to move away or exaggerated response	2
	Abn: Moves away after short period of time	1
	Abn: Does not move or reacts with excessively exaggerated response	0
Body condition score	1, emaciated; 2, thin; 3, normal; 4, overweight; 5, obese	1–5
Total score		1–13

Figure 2. Behavioral and appearance descriptions used to assign scores to mice. Abn, abnormal; N, normal.

**Provoked behavior.** After the assessment of natural behavior, each mouse was gently nudged and assigned a score of 3 to 0 representing its response to this provocation. A normal score of 3 was assigned to mice that readily walked or ran away or turned to sniff the observer. A score of 2 was given to mice that reluctantly moved away with some difficulty or had a hyperactive response inconsistent with mild provocation. A score of 1 was assigned to mice that moved slowly away after an extended pause. A score of 0 was assigned to animals that did not move or reacted with an especially exaggerated excitable response. Scores of 0 and 1 were considered abnormal.

**Body condition score and weight.** Mice were weighed every other day. For assessment of BCS, mice were placed on the wire-bar top of the cage. While gently restraining a mouse by holding the base of its tail, the observer used the thumb and index finger of the other hand to palpate the degree of muscle and fat over the sacroiliac region. Animals were scored every other day for body

condition on a scale from 1 to 5, according to previously established criteria.<sup>6,7,24</sup> Only whole numbers were assigned. Briefly, a mouse receiving a score of 1 was considered emaciated, with no palpable fat over the sacroiliac region, severely reduced muscle mass, with prominent vertebrae and iliac crests. A score of 2 was given to those animals with some fat deposition and muscle mass but less than that palpated in mice with a score of 3; these mice also had visible iliac crests. Mice that received a score of 3 were considered normal, had easily palpable fat pads, reduced definition of vertebral bodies, palpable but not visible iliac crests, and thick prominent muscle mass. Mice assigned a score of 4 were ‘overweight,’ as demonstrated by difficulty in palpating iliac crests, difficulty in assessing vertebral definition, and prominent fat pads overlying muscled areas. Obese mice were given a score of 5; these mice had fat pads that overlaid muscle and iliac crests, thereby obscuring their presence both tactilely and visually and giving the animal’s rump a rounded appearance. Body condition scores of 1 or 2 were considered abnormal.

**Total score.** Total scores were calculated by combining the sub-scores assigned after subjective evaluation of appearance, natural behavior, provoked behavior, and BCS. Weight was excluded from the total score because it was an objective measurement that varied depending on the age and initial size of the mouse. Total scores were calculated and ranged from 1 to 13, with 13 representing an obese animal with normal behavior and 1 representing a weak, hunched, nonambulatory, emaciated animal with an unkempt hair coat. Animals with a score of 11 to 13 were considered healthy, whereas animals with a score of 6 to 10 were considered less robust and demonstrating clinical signs associated with morbidity. When an animal’s total score fell at or below 5, the animal was scored and weighed daily by our study group, until a continued decline in score necessitated euthanasia according to the IACUC-approved endpoints of the protocol or the veterinarian’s discretion. We chose to increase the frequency of assessment to enhance our ability to advocate euthanasia if warranted. If the animals did not improve in total score after 2 d, the veterinarian was consulted. We found that animals with a score of 3 or less were candidates for euthanasia, especially if the BCS was 1, consistent with the IACUC-approved endpoints.

**Data analysis.** Of the 40 mice that started in the study, 36 were included in the data set for analysis. Of the 24 animals in the endpoint assessment experimental group, 20 were included in the data analysis. The remaining 4 mice were excluded because they either developed independent health problems (hydrocephaly, n = 1) or because they did not develop signs of tumor development (abdominal distention, n = 3) by the completion of the study. No animals that developed clinical signs consistent with tumor development remained at the 46-d study endpoint. All animals in



the control groups ( $n = 16$ ) were included in the data analysis and did not decline in health.

Our data analysis was performed without stratifying the endpoint assessment experimental groups into specific treatment type. Although we realize that the various treatments (antibody treatment versus antibody plus vaccine) may have altered the rate and extent of tumor growth, the purpose of our study was to assess BCS and behavior scoring, relative to a control group, as an adjunctive monitoring method. Analysis of the efficacy of the different tumor treatments will be reported elsewhere by our collaborator.

Body weight, BCS, appearance, natural behavior, provoked behavior, and total score, were compared between the endpoint assessment control and endpoint assessment experimental groups. These data were normalized for the variation in time each mouse was in the study before being euthanized, because animals were not evaluated for the same number of days. Weights were calculated as a percentage change from the starting to terminal weight divided by the number of days the mouse was in the study before being euthanized due to advanced tumor burden (according to the endpoints approved by the IACUC). The BCS, appearance, natural behavior, provoked behavior, and total score data were calculated as numerical score change from initial day of study to endpoint divided by the total number of days each particular mouse was in the study before euthanasia. Data are shown as mean  $\pm$  2 SD. Statistical differences were calculated by using an unpaired  $t$  test for the weight comparisons and both unpaired  $t$  and Mann–Whitney tests for the subjective scoring. All statistical analyses were performed by using SigmaStat statistical software (version 3.0, SPSS, Chicago, IL). Differences were considered significant if the  $P$  value was less than 0.05.

## Results

Two mice in the first antitumor antibody group (E1) and 1 mouse in group E3 did not develop clinical signs consistent with tumor development (lethargy and abdominal distention) after 46 d. These mice were excluded from the data analysis of the endpoint assessment study but were included in our collaborator's efficacy study. Their scores were consistent with those of the control groups. One mouse in the first antitumor antibody group (E1) developed hydrocephaly unrelated to study manipulations after study day 4 and was removed from both studies.

**Weight comparisons.** Body weights increased for all animals, with the control group weights increasing an average of  $0.46\% \pm 0.19\%$  of their initial weight each day and the experimental group increasing by an average of  $0.62\% \pm 0.62\%$  of their initial weight daily. The length of participation in the study varied for each mouse in the experimental group and ranged from 25 to 46 d (Table 1). Mice in the endpoint assessment control groups and experimental groups did not differ significantly in terms of changes in body weight from baseline values (Figure 3).

**BCS comparisons.** All animals in the study started with a normal BCS of 3 on a scale of 1 to 5. In the endpoint assessment control groups ( $n = 16$ ), 10 mice increased from a score of 3 to 4, correlating with the weight gain, whereas the remaining 6 mice maintained a BCS of 3. When the data from all animals in the endpoint assessment control groups were averaged, the daily change in BCS was  $0.014 \pm 0.022$  (mean  $\pm$  2 SD). In the endpoint assessment experimental groups, BCS decreased by an average of  $0.060 \pm 0.033$  daily. Of the 20 animals in the experimental groups, 18

received a BCS of 1 by the time they were removed from the experiment and euthanized (Table 1). The difference in BCS change between the control and experimental groups was statistically significant ( $P < 0.001$ ).

**Appearance.** The difference in daily average appearance score change for the endpoint assessment control and experimental groups was statistically significant ( $P < 0.001$ ). All mice in the control groups maintained a normal score of 2 on a scale of 0 to 2 throughout the study. Although animals in the experimental group started with a score of 2, most mice had decreased scores to 0 by the end of the study (Table 1). The average daily appearance score loss for the experimental groups was  $0.059 \pm 0.043$  (versus  $0 \pm 0$  for controls).

**Natural behavior.** The difference in average daily natural behavior scores for the endpoint assessment control and experimental groups was statistically significant ( $P < 0.001$ ). Natural behavior scores did not change for the control group: all scores remained at 3 on a scale of 0 to 3 throughout the study. All mice in the experimental group started at scores of 2 (4 mice) or 3 (16 mice) and decreased to 2, 1, or 0 before being removed from the study (Table 1). The average natural behavior score loss daily for the experimental groups was  $0.059 \pm 0.03$  (versus  $0 \pm 0$  for controls).

**Provoked behavior.** As with the appearance and natural behavior scores, the difference in the average daily provoked behavior scores for the endpoint assessment control and experimental groups was statistically significant ( $P < 0.001$ ). The control groups started and ended at 3 on a scale of 0 to 3, with no change. The experimental groups all started the experiment at a score 3 but ended at different scores (Table 1). The average score loss daily for the experimental groups was  $0.061 \pm 0.066$  (versus  $0 \pm 0$  for endpoint assessment controls).

**Total score.** All animals regardless of treatment group began the study with a total score of at least 10 of a maximum of 13. In the endpoint assessment control groups, the total score increased to 12 in 10 mice, including animals that had an increase in BCS from 3 to 4. The remaining 6 control animals maintained a total score of 11 until the end of the experiment. The average increase per day of TS for the endpoint assessment control groups was  $0.0136 \pm 0.022$ . The endpoint assessment experimental groups had an overall daily decrease in total score of  $0.24 \pm 0.15$  (Figure 4). Individual and total score changes are summarized in Table 1. By the completion of the study, 3 animals had total scores of 1, 1 animal had a total of 2, 8 mice had total scores of 3, 5 animals had total scores of 4, 1 animal had a score of 8, and the remaining 2 mice had total scores of 10. None of the mice in the experimental groups showed increases in TS. The difference between the average daily change in total score between the endpoint assessment control and experimental groups was statistically significant ( $P < 0.001$ ).

**Representative time-course assessment.** The preceding data were analyzed as numeric score difference between study beginning and end points divided by number of study days, which varied among mice. Tracking the same data over time for 1 representative animal from both the experimental group (Figure 5 A) and the control group (Figure 5 B) reveals trends in weight and total score as the disease process progresses. For the experimental group animal, weight increased whereas total score decreased. In contrast, in the control animal, weight increased and total score changed little.

**Table 1.** Initial and terminal data of individual mice in the endpoint assessment experimental group

Group	Mouse	Weight (g)		Body condition score		Appearance		Normal behavior		Provoked behavior		Total score		No. of days on study
		Initial	Terminal	Initial	Terminal	Initial	Terminal	Initial	Terminal	Initial	Terminal	Initial	Terminal	
E1	1A	14.5	19	3	1	2	2	3	2	3	3	11	8	31
	1D	16.2	18.6	3	1	2	0	3	1	3	2	11	4	29
	2A	15.3	21.4	3	1	2	0	3	0	3	0	11	1	42
	2B	15.9	17.6	3	1	2	0	3	1	3	1	11	3	31
	2C	16.6	20.3	3	1	2	0	3	1	3	2	11	4	36
	2D	17.3	20.1	3	1	2	0	3	1	3	2	11	4	31
	3B	16	19.3	3	1	2	0	3	1	3	1	11	3	30
	3C	16.8	21.4	3	3	2	2	3	2	3	3	11	10	46
	3D	14.9	16.8	3	1	2	0	3	1	3	0	11	2	30
E2	14A	17.5	19.1	3	1	2	0	2	0	3	0	10	1	28
	14B	16.1	18.8	3	1	2	0	2	0	3	0	10	1	28
	14C	15.6	21	3	1	2	0	2	1	3	1	10	3	30
	14D	17.6	21.5	3	2	2	0	2	1	3	1	10	4	25
	15A	15.9	17.2	3	1	2	0	3	1	3	0	11	2	26
	15B	17	21.8	3	1	2	0	3	1	3	1	11	3	28
	15C	16.9	19	3	1	2	0	3	1	3	2	11	4	35
	15D	16	21.4	3	1	2	0	3	1	3	1	11	3	33
E3	16B	16.5	18.5	3	1	2	0	3	1	3	1	11	3	32
	16C	15.2	15	3	1	2	0	3	1	3	1	11	3	35
	16D	16.8	20.6	3	1	2	0	3	1	3	1	11	3	33

All of these mice were challenged with tumor cells. In addition, the E1 group received antiVB6 chimeric tumor antibody, the E2 group received antiCD4 chimeric tumor antibody along with an idiotypic vaccine, and the E3 group received antiVB7 chimeric tumor antibody. The data for the endpoint assessment control groups was more consistent between animals and therefore are not shown.

## Discussion

Weight loss is a commonly used endpoint for animal studies,<sup>11,15,23,25</sup> but is not an ideal parameter for monitoring animal wellbeing in studies involving internal tumor growth. The increase in the mass of the tumor can mask the loss of overall body condition and dehydration, thereby delaying interventions such as parenteral or nutritional support and euthanasia. For subcutaneous tumors, reported alternative endpoints include tumor size, presence or absence of ulceration, and interference with mobility.<sup>8</sup> Abdominal tumors are more challenging, because it is difficult to assess tumor size without complex imaging techniques (for example, computed tomography and magnetic resonance imaging) or distress to the animal (for example, restraint and palpation). Body condition score may aid in health assessment because it is independent of body weight and abdomen size. Instead, BCS allows assessment of muscle mass and adipose accumulation, thereby monitoring cachexia and emaciation.

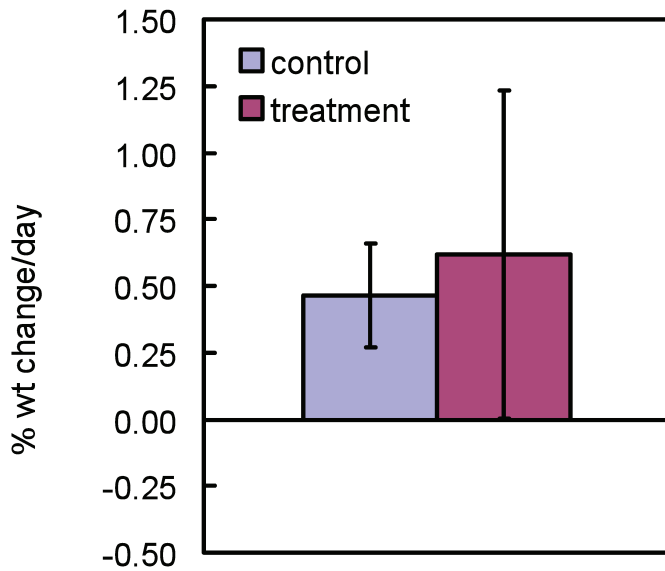
Our study intended to assess the applicability of a variety of parameters that can be easily and regularly assessed with minimal training and equipment. These parameters included body weight, BCS, appearance, natural behavior, and provoked behavior. Appearance and behavior are commonly assessed markers of animal wellbeing. However, these criteria are highly subjective, unless observers are asked to use a scoring system to classify each animal.<sup>11,15</sup> To develop an assessment scale, we evaluated these parameters individually and as a group (that is, total score), excluding body weight.

In this study, all animals gained weight regardless of treatment group. This result was understandable, because the subjects, being young female mice, should have gained weight as they continued to grow. However, we expected that as the experimental groups became ill and their tumor burden increased, their weight would either increase or remain unchanged. Our data supported this prediction and demonstrated that weight was not an accurate indicator of the morbidity of tumor burden. As demonstrated by the time courses of total score and body weight for a representative mouse (Figure 5), an animal's weight can increase significantly while the total score, reflecting overall health, declines rapidly.

The mean rate of weight change of the experimental groups was not statistically significant different than that of the control group. Because the mice in the experimental groups reached endpoints earlier and were therefore on study for a shorter period of time than were the control mice (25 to 46 d for experimental mice versus 46 days for all mice in the control group; Table 1), the rate of weight increase between groups may have diverged more with time had all of the experimental mice completed the full 46 d of the study. Abdominal distension related to tumor growth may have masked some differences.

Compared with body weight, BCS was a more accurate indicator of mouse health and decreased as mice began to develop clinical signs associated with tumor growth (for example, abdominal distention, reduced activity). This result was expected because BCS is assessed by evaluating muscle mass and fat deposition in the sacroiliac region, which decrease as neoplasia progresses and

## Weight



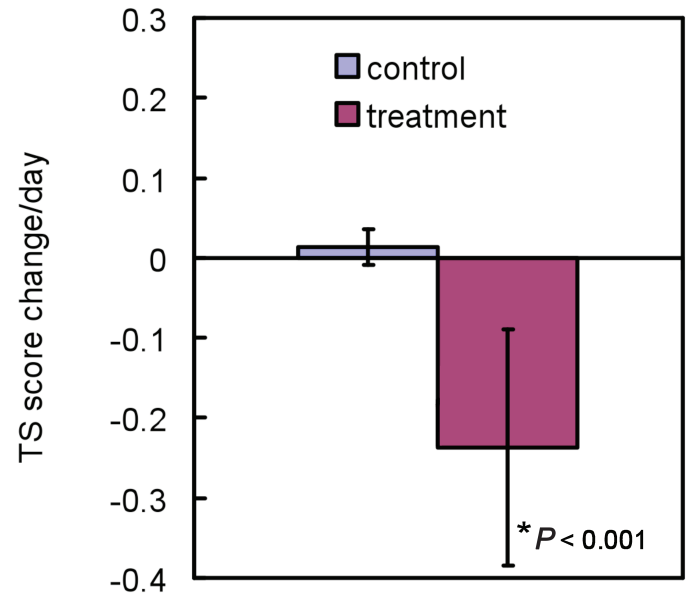
**Figure 3.** Percentage of weight change from baseline daily in endpoint assessment control (n = 16) and experimental (n = 20) groups. Each mouse's percentage of weight change from terminal to baseline was divided by the number of days that that mouse participated in the study. The average of these data was calculated for the endpoint assessment control and experimental groups. Error bar, 2 SD.

the animal develops cachexia and becomes ill. Because it is independent of body weight, BCS is a more accurate reflection of the health of an animal. This measurement generally is accepted in companion species as an assessment of nutritional status. Possible complications with this monitoring technique may arise if subcutaneous edema or fatty tumor development over the sacroiliac region is present; but these complications likely would be rare in young rodents over a short (46-d) study period.

The appearance score was also more reliable than body weight as an indicator of health for the mice in this study. All of the animals in the control groups maintained good grooming habits, whereas those in the experimental groups developed piloerection and decreased grooming as the tumors increased in size. Although poor grooming may not be a direct indicator of tumor burden, it may reflect the mouse's activity level and health status. However, because this scoring system only ranged from 0 to 2, scores showed little variability, even in the experimental groups. This narrow scoring system might not detect subtle changes that would otherwise be detected in total score calculations.

The behavioral scores were more reliable than body weight in indicating when a mouse was ill. Behavior can be an indication of energy level: if animals are healthy, they typically are highly active. However, this assessment depends on a mouse's natural activity periods and other factors, such as stress (previously handled or manipulated, newly cleaned cage). The natural and provoked behavior scores both were ranked from 0 to 3, resulting in the possibility of more variation in scores than seen with appearance scoring. The provoked scores in the experimental groups were more varied than the appearance or natural behavior scores and were anywhere from 0 to 3 by the completion of the

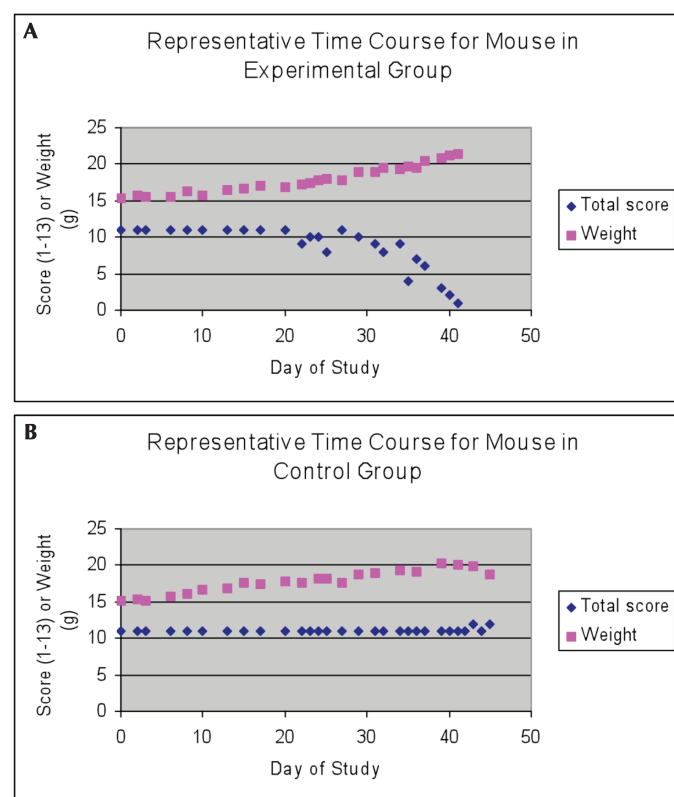
## Total Score



**Figure 4.** Change in total score of endpoint assessment control (n = 16) and experimental (n = 20) groups as averaged daily. Data were calculated as numerical score change from initial day of study to endpoint divided by the total number of days each mouse participated in the study before euthanasia. Error bars, 2 SD.

experiment. This result suggests that provoked behavior scores should not be used as the sole measure and encourages the use of total score calculation to most accurately assess animal health. Because these scores are subjective measurements, they likely reflect variability between observers and mouse responses. Depending on the time of day, amount of stress, and other factors, each mouse may respond differently to a standard provocation at various times. The mice in this experiment were assessed at the same time during the day, but because mice are nocturnal animals, consideration could be given to assessing behavior scores at night in a future study. In addition, although one can reasonably expect a healthy animal to have enough energy to ambulate about the cage, sometimes bursts of activity can occur in injured or sick animals, as in a fight-or-flight response, which supports why a hyperactive response would also have received a score of 0. In this study, no animal exhibited a hyperactive response.

Although the BCS, appearance, natural behavior, and provoked behavior scores are all helpful indicators of morbidity in this abdominal tumor model, they combine to produce the total score, which is the most collective and sensitive assessment likely to indicate poor health. Each individual parameter contributes to the overall total score, but because each parameter is also subjective, their combined assessments more accurately illustrate the picture of animal health in this tumor model. In this study, we found that when an animal's total score fell below 5 (maximum, 13), increases in the frequency of evaluation were indicated to allow timely identification of endpoints. A total score of 3 or less generally indicated that euthanasia was appropriate. An animal with a score of 3 typically had a stable body weight but appeared thin, did not groom, and moved little, despite provocation. In



**Figure 5.** Time course for a representative mouse from the (A) experimental and (B) control groups. Note that despite its gradual weight gain, the experimental mouse's (A) total score decreases. The decrease in total score remains constant until it abruptly decreases. The control mouse's (B) weight gradually increases as its total score remains unchanged.

traditional rodent studies, thin appearance, a decrease in grooming, or lethargy may not warrant euthanasia according to IACUC-approved endpoints. By collectively evaluating all of these points as well as behavior scores, we were able to develop a more complete clinical picture and identify health deterioration sooner than was possible with our traditional IACUC-approved endpoints. Although some animals with scores below 3 did not necessarily fulfill all of the traditional IACUC-approved endpoints for euthanasia, both our observers and our collaborator's staff had concerns for these animals that would normally cause them to seek veterinary intervention. This outcome illustrates a weakness in the IACUC-approved subjective criteria and how application of objective ratings can improve the welfare of the animals on a study. Because mice with a total score of 5 consistently fell within 72 h to a total score of 3 (consistent with the subjective IACUC-approved endpoints), we recommend using this objective rating scale with increased monitoring frequency when mice reach a total score of 7, with euthanasia when they reach a total score of 5.

In traditional IACUC-approved protocols, monitoring parameters have included weight, appearance, and behavior. Appearance and behavior are subjective criteria and can vary markedly between observers, a considerable drawback relative to the more commonly used endpoint assessments. Because these parameters are subjective, emphasis typically is placed on changes in weight as a reflection of anorexia or dehydration, indicative of declining animal health.<sup>11,15</sup> However, body weight may remain stable

in studies involving tumor growth, and assessment of behavior becomes more difficult as a sole indicator for euthanasia. In our study, total score was more accurate for assessing mouse health in these abdominal lymphoma studies than was the use of any single criterion.

In this abdominal tumor model, abdominal distention indicated tumor development, and although present, it was not quantified according to a scale. Because these mice were on another investigator's study, necropsies were not performed after removal from the study to ascertain tumor burden or ascites. This information would especially have been helpful in the 3 experimental animals that were removed from the data set due to lack of onset of clinical signs associated with tumor development or abdominal distention.

The application of BCS in abdominal tumor models does have limitations, which necessitate additional research on this method as a monitoring parameter. For example, animal age, gender, and reproductive status become factors in BCS and total score assessment. As animals age, muscle mass decreases and fat deposition shifts.<sup>5</sup> Pregnancy also can alter BCS assessment. If similar studies are performed in an aged population, the observer may assign a low BCS and total score to animals that are still healthy for their age. The scale may need to be modified for studies involving aged or reproductively active animals.

Another area of further study involves assessment of the time point at which the total score first indicates ill-health. As illustrated by the time course of a representative mouse (Figure 5) in this study, once total score decreased, it declined rapidly. Additional research on other models may reveal different rates of decline in total score. Therefore, further assessment of other models may assist animal care staff in monitoring animals every few days initially and then more frequently once the total score starts to decrease. For the present model, the monitoring intervals we described were appropriate.

Although the total score panel (including BCS and behavior assessments) presents a learning curve and requires manual labor and a time commitment to perform, it is easy to learn, is noninvasive, and can be performed by anyone, regardless of education and skill level. In addition, obtaining the total score is relatively resistant to interobserver variability.<sup>8</sup> Body weights alone can be variable, due to differences between equipment, time of day, and personnel. If total score assessment is done correctly, the time involved may not greatly exceed that for obtaining body weights only, and the data will be more reliable. Additional studies can be performed to assess time commitments for various morbidity assessment techniques.

Assessing BCS, appearance, and behavior are appropriate adjunct assessments to weight monitoring in abdominal lymphoma models, for which visual assessment of tumors is not possible. In addition, these criteria potentially could serve as adjunct monitoring parameters in other mouse models in which body weight is not an accurate indication of health. Such models include other abdominal tumors, cutaneous tumors, surgical device implantation, infectious disease, organ hypertrophy, and cardiovascular disease. BCS and total score assessments are helpful methods of health evaluation in mice and can be considered adjunct monitoring techniques in protocols.



## Acknowledgments

The authors would like to thank Craig Okada for allowing the concurrent collaboration with his study animals. They would also like to thank Carla Webb and Rachel Luksic for their assistance in data collection, David Mailett for his assistance in data illustration, and Stephanie J Murphy for her assistance in data analysis and manuscript review. This study was funded in part by grants from the John Hopkins Center for Alternatives to Animal Testing (CAAT) and the Portland Veterans Affairs Research Foundation (PVARF).

## References

1. Aldred AJ, Cha MC, Meckling-Gill KA. 2002. Determination of a humane endpoint in the L1210 model of murine leukemia. *Contemp Top Lab Anim Sci* 41:24–27.
2. Baez JL, Michel KE, Sorenmo K, Shofer FS. 2007. A prospective investigation of the prevalence and prognostic significance of weight loss and changes in body condition in feline cancer patients. *J Feline Med Surg* 9:411–417.
3. Cartmill JA, Thompson DL, Storer WA, Gentry LR, Huff NK. 2003. Endocrine responses in mares and geldings with high body condition scores grouped by high versus low resting leptin concentrations. *J Anim Sci* 81:2311–2321.
4. Christie JL, Hewson CJ, Riley CB, McNiven MA, Dohoo IR, Bate LA. 2006. Management factors affecting stereotypies and body condition score in nonracing horses in Prince Edward Island. *Can Vet J* 47:136–143.
5. Das M, Gabriely I, Barzilai N. 2004. Caloric restriction, body fat, and aging in experimental models. *Obes Rev* 5:13–19.
6. Easterly ME, Foltz CJ, Paulus MJ. 2001. Body condition scoring: comparing newly trained scorers and microcomputed tomography imaging. *Lab Anim*. 30:46–49.
7. Foltz CJ, Ullman-Cullere M. 1999. Guidelines for assessing the health and condition of mice. *Lab Anim* 28:28–32.
8. Hickman DL. 2007. Use of body condition scoring as an adjunct endpoint for tumor growth studies. *J Am Assoc Lab Anim Sci* 46:111.
9. Hoedemaker M, Prange D, Gundelach Y. 2008. Body condition change ante- and postpartum, health and reproductive performance in German Holstein cows. *Reprod Domest Anim Epub ahead of print*.
10. Hoybergs YM, Biermans RL, Meert TF. 2008. The impact of bodyweight and body condition on behavioral testing for painful diabetic neuropathy in the streptozotocin rat model. *Neurosci Lett* 436:13–18.
11. Jones HRP, Oates J, Trussell BA. 1998. An applied approach to the assessment of severity, p 40–47. In: Hendriksen CFM, Morton DB, editors. *Humane endpoints in animal experiments for biomedical research*. London (UK): Royal Society of Medicine Press.
12. Laflamme DP. 2005. Nutrition for aging cats and dogs and the importance of body condition. *Vet Clin North Am Small Anim Pract* 35:713–742.
13. Laflamme DP. 2006. Understanding and managing obesity in dogs and cats. *Vet Clin North Am Small Anim Pract* 36:1283–1295.
14. Lents CA, White FJ, Ciccioli NH, Wettemann RP, Spicer LJ, Lalman DL. 2008. Effects of body condition score at parturition and postpartum protein supplementation on estrous behavior and size of the dominant follicle in beef cows. *J Anim Sci* 86:2549–2556.
15. Lloyd MH, Wolfensohn SE. 1998. Practical use of distress scoring systems in the application of humane endpoints. p 48–53. In: Hendriksen CFM, Morton DB, editors. *Humane endpoints in animal experiments for biomedical research*. London (UK): Royal Society of Medicine Press.
16. Morton DB. 2000. A systematic approach for establishing humane endpoints. *ILAR J* 41:80–86.
17. Richmond J. 1998. Criteria for humane endpoints. p 26–32. In: Hendriksen CFM, Morton DB, editors. *Humane endpoints in animal experiments for biomedical research*. London (UK): Royal Society of Medicine Press.
18. Salak-Johnson JL, Niekamp SR, Rodriguez-Zas SL, Ellis M, Curtis SE. 2007. Space allowance for dry, pregnant sows in pens: body condition, skin lesions, and performance. *J Anim Sci* 85:1758–1769.
19. Schiffer SP. 1997. Animal welfare and colony management in cancer research. *Breast Cancer Res Treat* 46:313–331.
20. Slupe JL, Freeman LM, Rush JE. 2008. Association of body weight and body condition with survival in dogs with heart failure. *J Vet Intern Med* 22:561–565.
21. Stasiak KL, Maul D, French E, Hellyer PW, VandeWoude S. 2003. Species-specific assessment of pain in laboratory animals. *Contemp Top Lab Anim Sci* 42:13–20.
22. Tolkamp BJ, Emmans GC, Kyriazakis I. 2006. Body fatness affects feed intake of sheep at a given body weight. *J Anim Sci* 84:1778–1789.
23. Toth LA. 2000. Defining the moribund condition as an experimental endpoint for animal research. *ILAR J* 41:72–79.
24. Ullman-Cullere MH, Foltz CJ. 1999. Body condition scoring: a rapid and accurate method for assessing health status in mice. *Lab Anim Sci* 49:319–323.
25. Wallace J. 2000. Humane endpoints and cancer research. *ILAR J* 41:87–93.